



# Integrating TEI and EAD

Chris Turner  
LEADERS Project,  
School of Library, Archive and Information Studies,  
University College London

Copyright, UCL

LEADERS: Linking EAD to Electronically  
Retrievable Sources



# Linking EAD to Electronically Retrievable Sources

- Developing a generic computer-based toolset that integrates EAD encoded finding aids with TEI encoded transcripts and digitised images of archival material.
- Text Encoding Initiative (TEI)
  - For archival source materials
- Encoded Archival Description (EAD)
  - For finding aids
- NISO Metadata for Images in XML (MIX) Schema
  - For digital images metadata



# TEI for Archives: Research Methodology

- Analysis of commonly occurring structures, features and contents found within a range of different types of archive source material
  - Material held in UCL's Special Collections and Record Office
  - Expect to analyse material held in other archival repositories to validate initial findings and uncover as wide range of encoding challenges as possible



# TEI for Archives: Preliminary findings

- ‘Transcription of Primary Sources’ tag set in TEI can deal with a wide range of encoding challenges inherent in archival material:
  - Complex additions, deletions and corrections
  - Gaps within and damage to the text
  - Changes in document hands, style and character of writing



# TEI for Archives: Preliminary findings

- Need to explore encoding options for:
  - ‘overlaid data’
  - Textual and numerical data presented in complex tables
  - Formulae and mathematical expressions within the text
- Examining TEI and other DTDs specifically built to handle these structures and features



## Exploring solutions for encoding 'overlaid data'

- 'overlaid data': when an underlying layer of data is used as the basic structure onto which further data [other layer(s)] is applied
  - Accounts and registers
  - Address books
  - Calendars
  - Questionnaires and forms



# Example of 'overlaid data'

UNIVERSITY COLLEGE, LONDON.

Form to be filled up by every person wishing to become a Student of the College (so far as it may apply to his or her case).

Name in full Marie Babut

Address while attending the College 4 Brundrick Square W.C.

Usual Address 1 rue Boudaloue, Primes, Gard-France

Date of Birth 28 November 1873.

Previous place of education Primes.

Public Examinations (if any) already passed \_\_\_\_\_

Name of Parent or Guardian Monsieur P. Babut

Address of Parent or Guardian 1 rue Boudaloue Primes

Any Examination which the intending Student has in view should be indicated here by putting a cross (X) below the name of it in the following list, if it occurs therein; otherwise the name of the Examination should be inserted in writing.

EXAMINATIONS OF THE UNIVERSITY OF LONDON.

Matriculation.	Interm. Arts.	Interm. Sci. Medical.	Profin. Sci. Medical.	Interm. Medical.	B.A.	D.Sc.	M.B.	M.A.	D.Sc.

Examinations other than the above:

Work at the Trade School.

Declaration of Adhesion to Regulations.

I, the undersigned, do hereby engage that I will conform to such Regulations as have been, or may be, made for the maintenance of order in the College and in the Classes which I attend.

Date 14<sup>th</sup> October 1896

(Signed) M Babut *MA Morrison*



# TEI/EAD Integration: Overlaps

- Overlaps between EAD and TEI occur in relation to metadata that:
  - Identifies, locates and describes the creation of the original archive document
  - Describes the physical characteristics of the original object
  - Provides contextual information about the creator of the original object and the participants within the object
  - Interprets/describes the data in the object





# TEI/EAD Integration: Objectives

- Avoidance of repetition of information within EAD and TEI
- More complete description of an object's physical characteristics
- Standardisation of contextual information regarding individuals/organisations



# TEI/EAD Integration: search and retrieval issues

- Elements that interpret/categorise the actual data within the object

EAD Elements		Overlapping TEI elements	
Name of immediate parent element	Name of child elements	Name of immediate parent element(s)	Name of child elements
<controlaccess>	<genreform> <geogname> <persname> <famname> <corpname> <occupation> <subject> <date> <function>	<profileDesc> within <teiHeader>	<keywords><classcode> classref
		<textDesc> within <profileDesc> within <teiHeader>	<channel> <constitution> <domain> <factuality> <preparadness> <purpose>
N/A	<scopecontent>		
		<text>	<name><date>



# LEADERS toolset

- XML Schema
  - Use of namespaces
  - Schema will provide a generic and re-usable means to encode resources
- XSLT/CSS
  - Style sheets will provide a means to manipulate, transform and present the encoded resources, thus supporting re-purposing of encoded materials
- WSDL/SOAP
  - Incorporating ‘self-describing services’ will allow multiple applications to be constructed to make different use of the encoded materials



# Sample application

- Demonstrator – a sample application to show what can be produced/generated from the encoded materials and the toolset
- Basic search and retrieval and alternative presentations to show the possibilities of TEI/EAD encoded resources
- Used to gather feedback from users which will guide us in further design and development